Bioinformatics tools in molecular oncology

Jasmina M. Obradovic[1*], Branko J. Arsic[2]

[1] Institute for Information Technologies, Department of Science, University of Kragujevac, Kragujevac, Republic of Serbia; e-mail: jasmina.obradovic@uni.kg.ac.rs
[2] Faculty of Science, University of Kragujevac, Kragujevac, Republic of Serbia; e-mail: branko.arsic@pmf.kg.ac.rs

* *Corresponding author*

DOI: 10.46793/ICCBIKG25.323O

Abstract: Bioinformatics and molecular oncology are rapidly evolving scientific field, so their integration has become fundamental, offering powerful capabilities for mutation detection, driver gene identification, prognostic analysis and personalized treatment strategies. In this brief descriptive review, several key bioinformatic tools are presented, within the constraints of the paper's format, highlighting their advantages and potential limitations. These tools and wide array of publically available databases and platforms support research in this area. However, challenges such as data quality, computational infrastructure, adequate expertise, and the need for both experimental and clinical validation must be addressed in the future studies. Nevertheless, the continuous advancements of bioinformatics tools is expected to improve the concept of personalized medicine and potentially lead to improved patient outcomes.

Keywords: bioinformatic tools, molecular oncology, multiomics

1. Introduction

In the era of multiomics, bioinformatic tools have become indispensable in molecular oncology, enabling researchers and clinicians to analyze, visualize, and interpret vast and complex molecular data related to cancer [1]. In this mini-review we provide a concise overview of this complex, rapidly-evolving and highly promising field, along with key bioinformatic tools and approaches commonly used in molecular oncology.

2. Methodlogy

In this narrative review, we analyzed literature published over the past 20 years. For this purpose, manuscripts were obtained from publicly available databases such as PubMed and Scopus. We identified a broad range of bioinformatic tools with applications including gene expression analysis and variant discovery—such as single nucleotide polymorphisms (SNPs), mutations, and copy number variation (CNV) analysis. Additional applications include DNA methylation profiling, chromatin accessibility analysis, epigenetic regulation, histone modification analysis, protein identification via mass spectrometry, and visualization of protein interactions. Furthermore, we identified

tools for RNA sequencing (RNA-seq), noncoding RNA analysis, and single-cell RNA sequencing (scRNA-seq). However, due to space constraints, only a selection of these tools is highlighted in this review.

3. Results and Discussion

High-throughput molecular technologies like next generation sequencing (NGS), RNA-sequencing, and mass spectrometry provide identification of novel mutations, potential biomarkers and therapeutic targets [2]. The quality of raw data obtained from sequnecing methods is essential for accurate bioinformatic analyses. Preprocessing challenges may arise during library preparation, thus, sample contamination, sequencing errors, and batch effect could introduce certain bias, thus leading to incorrect conclusions. Tools like Genome Analysis Toolkit (GATK) are highly effective in mutation detection (germline or somatic), as well as for identifying insertions/deletions (indels), or single nucleotide polymorphisms (SNPs), although sensitive to sequencing errors [3]. A component of GATK, Mutest2, is useful in somatic mutation detection. Mutest2 and VarScan could detect point mutations, rare or structural variations that might be overlooked with traditional methods. These capabilities are valuable in identification of less common cancer subtypes, that respond to other therapy regiments. VarScan, Strelka2, DESeq2, are used in identifying variants in DNA and RNA sequencing datasets [4-6].

Personalized treatment approach, (where right patient gets right therapeutic, at right time, and right dose), is reshaped with bioinformatic tools. Based on the individuals' genetic profile, these tools can help to identify specific mutations (e.g., *KRAS*, *EGFR*), or expression patterns, and in further determine the most effective therapies for a specific patient. Beyond the tools themselves, several interactive cancer genomics platforms such as cBioPortal, OncoKB, UCSC Genome Browser, or large datasets like The Cancer Genome Atlas (TCGA) allow comprehensive analysis of mutations, gene expressions data, and clinical parameters [7-10]. Although, certain caution could be addressed to these datasets in preventing generalisability, since TCGA mainly includes patients of European ancestry [11].

The range of "omics" now includes genomics, transcriptomics, proteomics, and extends further to epigenomics, metabolomics, radiomics, pathomics, microbiomics, and immunomics [1]. Tools that might highlight cancer progression by integrating multi-omics data are iClusterPlus and OmicNavigator [12, 13]. Majority challenging steps arise in this field, one of that might be large-scale input data, but besides adequate processing of complex data by bioinformatic tools, the interpretation of results requires careful and accurate validation, and specific expertise. A lack of standardization in bioinformatics workflow in oncology, as well as insufficient collaborations among clinicians, bioinformaticians and researchers, has been frequently cited as a source of bias and inefficiency [14].

Earlier-generation tools, such as Oncodrive or DriverDB exome database, are valuable in identification of genes with highest likelihood of being true cancer driver

genes, and thus help in identification of potential therapeutic targets [15, 16]. Many of those tools depend on tumor heterogeneity and the quality of available data, that could affect their accuracy. On the other side, strict computational algorithms, might not be true reflection of biological processes in cancer, so for predictions about mutations, drug responses, and generally for their applicability in clinical practice, experimental validation is mandatory.

Patient stratification into risk categories, treatment planning, and correlation of gene expression profiles with patient survival outcomes is enabled by tools like GEPIA, KM Plotter and SurvExpress [17-19]. Furthermore, platforms and tools like DrugBank and Comparative Toxicogenomic Database (CTD) can be used to analyze drug-target interactions and can predict efficacy of the existing therapies for specific cancer-related mutations [20, 21]. A core issue in this field is cost-effectiveness. Bioinformatic analyses of large datasets requires substantial computational infrastructure and specialized expertise. An example of this is PathAI, a deep learning approaches for cancer image analysis, illustrating an increasing role of artificial intelligence in diagnostic and prognostic procedures [22].

4. Conclusions

In conclusion, in molecular oncology, bioinformatic tools offer majority of opportunities for advancing cancer research and precision medicine, enabling the identification of novel biomarkers, cancer subtypes, and target therapies. Although, current challenges (data quality, lack of standardization, and need for clinical validation) remain, but many of these obstacles are expected to be overcome in the near future.

Acknowledgment

This work is funded by the Ministry of Science, Technological Development and Innovation, Republic of Serbia, Agreements No. 451-03-136/2025-03/200378 and 451-03-137/2025-03/200122.

References

[1] C. Chen, J. Wang, D. Pan, et al., *Applications of multi-omics analysis in human diseases*. MedComm, 4(4), (2023) e315.

[2] D. Rosati, M. Palmieri, G. Brunelli, et al., *Differential Gene Expression Analysis Pipelines and Bioinformatic Tools for the Identification of Specific Biomarkers: A Review*, Computational and Structural Biotechnology Journal 23 (2024) 1154-1168.

[3] Y. Zhou, N. Kathiresan, Z. Yu, et al., *A high-performance computational workflow to accelerate GATK SNP detection across a 25-genome dataset*. BMC Biol 22 (2024) 13.

[4] D. Benjamin, T. Sato, K. Cibulskis, et al., *Calling somatic SNVs and indels with Mutect2*. BioRxiv, (2019) 861054.

[5] Z. Chen, Y. Yuan, X. Chen, et al., *Systematic comparison of somatic variant calling performance among different sequencing depth and mutation frequency*. Scientific reports, 10(1), (2020) 3501.

[6] D. Li, M.S. Zand, T.D. Dye, et al.,. *An evaluation of RNA-seq differential analysis methods.* PloS one, 17(9), (2022) e0264246.
[7] P. Brlek, A. Kafka, A. Bukovac, et al., *Integrative cBioPortal Analysis Revealed Molecular Mechanisms That Regulate EGFR-PI3KAKT-mTOR Pathway in Diffuse Gliomas of the Brain*, Cancers 13, 13 (2021) 3247.
[8] J. Gao, B.A. Aksoy, U. Dogrusoz,G. et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal*. Sci Signal, 6 (2013) 1.
[9] D. Chakravarty, J. Gao, S. M. Phillips, et al., *OncoKB: A Precision Oncology Knowledge Base*, JCO Precision Oncology 2017 (2017): PO.17.00011.
[10] G. Perez, G.P. Barber, A. Benet-Pages, et al. *The UCSC Genome Browser database: 2025 update*. Nucleic acids research, 53(D1), (2025) D1243–D1249.
[11] Cancer Genome Atlas Research Network, J.N. Weinstein, E.A. Collisson, G.B. Mills, K.R. Shaw, B.A. Ozenberger, K.Ellrott, I. Shmulevich, C. Sander, J.M. Stuart. *The Cancer Genome Atlas Pan-Cancer analysis project*. Nature genetics, 45(10), (2013) 1113–1120.
[12] P. Chalise, D. Kwon, B.L. Fridley, Q. Mo.,. Statistical methods for integrative clustering of multi-omics data. In Statistical Genomics (pp. 73-93). New York, NY: Springer US. (2023)
[13] T.R. Ernst, J.D. Blischak, , P. Nordlund, et al. *OmicNavigator: open-source software for the exploration, visualization, and archival of omic studies*. BMC Bioinformatics, 25 (2024) 162.
[14] T. Wolde, V. Bhardwaj, V. Pandey., *Current Bioinformatics Tools in Precision Oncology*. MedComm, 6(7) (2025) e70243.
[15] D. Tamborero, N. Lopez-Bigas, A. Gonzalez-Perez et al., Oncodrive-CIS: a method to reveal likely driver genes based on the impact of their copy number changes on expression. PLoS One, 8(2), (2013) e55489.
[16] W.C. Cheng, I.F. Chung, C.Y. Chen, et al., *DriverDB: an exome sequencing database for cancer driver gene identification*. Nucleic acids research, 42(D1), (2014) D1048-D1054.
[17] Z. Tang, C. Li, B. Kang, et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses*. Nucleic acids research, 45(W1), (2017) W98-W102.
[18] B. Jia, X. Zhao, Y. Wang, et al., *Prognostic roles of MAGE family members in breast cancer based on KM-Plotter Data*. Oncology letters, 18(4), (2019) 3501-3516.
[19] R. Aguirre-Gamboa, H. Gomez-Rueda, E. Martínez-Ledesma, et al., *SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis*. PloS one, 8(9), (2013) e74250.
[20] C. Knox, M. Wilson, C.M. Klinger et al., *DrugBank 6.0: the DrugBank knowledgebase for 2024.* Nucleic acids research, 52(D1), (2024)D1265-D1275.
[21] A.P. Davis, T.C. Wiegers, D. Sciaky, et al., *Comparative toxicogenomics database's 20th anniversary: update 2025*. Nucleic acids research, 53(D1), (2025) D1328-D1334.
[22] B. Glass, M.E. Vandenberghe, S.T. Chavali, et al. *Deployment of a Machine Learning Algorithm in a Real-World Cohort for Quality Control Monitoring of Human Epidermal Growth Factor-2–Stained Clinical Specimens in Breast Cancer*. Arch Pathol Lab Med; 149 (8) (2025) 751–760.